Пример 1.1. Проверка соответствия выборки (данные в таблице справа) нормальному закону распределения при уровне значимости $\alpha = 0{,}05$.

На рис. А приведен скриншот MS Excel анализа нулевой гипотезы соответствия конкретной выборки непрерывных данных нормальному закону распределения.

При построении решения алгоритм предусматривает

- разбиения исходной выборки на определенное число классов (диапазонов), определяемых границами $x_{нач}$ $x_{кон}$;
- расчет экспериментальных частот n данной выборки;
- расчет теоретических (нормального закона распределения) частот $n^{теор}$ выборки такого же объема;
  Для нормального закона распределения
  $$n_i^{теор} = n\left[\Phi(x_i^{кон}) - \Phi(x_i^{нач})\right], \quad \Phi(x) = \int_{-\infty}^{z} \frac{1}{\sigma\sqrt{2\pi}} exp\left(-\frac{(x-\overline{x})^2}{2\sigma^2}\right) dx$$
  где $x_i^{нач}$, $x_i^{кон}$ – начало и конец i-того интервала,
- расчет $\chi^2_{экс} = \sum_{i=1}^{M} \frac{(n_i - n_i^{теор})^2}{n_i^{теор}}$ определение $\chi^2_{крит}$, их сравнение;
- анализ нулевой гипотезы посредством критерия χ^2.

9,0	47,5	63,0
17,5	48,0	64,5
21,0	50,0	65,0
26,5	51,0	67,5
27,5	53,5	68,5
31,0	55,0	70,0
32,5	56,0	72,5
34,0	56,0	77,5
36,0	56,5	81,0
36,5	57,5	82,5
39,0	58,0	90,0
40,0	59,0	96,0
41,0	59,0	101,5
42,5	60,0	117,5
43,0	61,0	127,5
45,0	61,5	130,0
46,0	62,0	

1. В диапазон A4:C20 заносятся исходные данные по выборке. Для удобства данный диапазону присваивается имя X – сначала диапазон выделяется, далее ВСТАВКА – ИМЯ – ПРИСВОИТЬ) {ВСТАВКА – НАЗВАНИЯ – ОПРЕДЕЛИТЬ}.
2. В ячейки E4:E8, H4:H7, E10:I10 и H19:H20 заносятся поясняющие данные.
3. В ячейку F8 заносится величина уровня значимости. Далее заполняются формулами следующие ячейки.

адрес	формула	пояснение
F4	=СЧЁТ(X) {=COUNT(X)}	подсчет n – количества элементов выборки
F6	=СРЗНАЧ(X) {=AVERAGE(X)}	расчет среднеарифметического значения $\overline{x}$
F7	=СТАНДОТКЛОН(X) {=STDEV(X)}	расчет стандартного отклонения σ
I4	=ОКРВВЕРХ(1+3,3*LOG10(F4);1) {=CEILING(1+3,3*LOG10(F4);1;1)}	подсчет M – количества классов (диапазонов) по формуле Sturgess'a $\mathrm{M} = 1 + 3{,}3 \cdot \lg n$
I5 I6	=МИН(X) {=MIN(X)} =МАКС(X) {=MAX(X)}	расчет x_{min} – минимального и x_{max} – максимального в выборке
I7	=ОКРУГЛВВЕРХ((I6-I5)/I4;1) {=ROUNDUP((I6-I5)/I4;1)}	расчет (округленной) ширины диапазона $\Delta x/M$

4. В ячейку E11 заносится формула =I5 – начальное значение первого диапазона, равное минимальному в выборке. В ячейку E12 заносится формула =E11+$I$7 – начальное значение второго диапазона, равное началу предыдущего плюс ширина диапазона.

Далее по содержимому E12 производится автозаполнение ячеек E13:E17 начальных значений для всех остальных диапазонов (интервалов, классов).

5. В ячейку F11 заносится формула =E11+$I$7 – конечное значение первого диапазона, равное его началу плюс ширина диапазона. Далее по F11 производится автозаполнение ячеек F12:F17 конечных значений для всех остальных диапазонов (классов).

6. В ячейки G11:G17 механизмом введения формул для массивов (использование F2 затем Ctrl+Shift+Enter) заносится формула =ЧАСТОТА(X; F11:F17) {=FREQUENCY(X;F11:F17), формула вводится в выделенные ячейки и активируется через Ctrl+Shift+Enter }. На этом этапе определяются количества элементов выборки (частота), относящиеся к каждому классу.

7. В ячейку H11 заносится формула =$F$4*НОРМРАСП(F11; $F$6; $F$7; ИСТИНА) {=$F$4*NORMDIST (F11; $F$6; $F$7;TRUE())}, которая определяет для нормального закона распределения (при соответствующих значениях $\bar{x}$ и σ) теоретически ожидаемое число элементов выборки для отрезка изменения переменной от -∞ до конца первого диапазона данных (см. формулу для $n_i^{\text{теор}}$). При этом используется встроенная в Excel функция НОРМРАСП() {NORMDIST()}, возвращающая стандартное нормальное *интегральное* распределение.

	A	B	C	D	E	F	G	H	I
1	**Нулевая гипотеза:** выборка ***X*** соответствует								
2	нормальному распределению								
3	выборка ***X***								
4	9,0	47,5	63,0		***n*** **=**	50		*M* =	7
5	17,5	48,0	64,5					***X*** *min* =	9,0
6	21,0	50,0	65,0		$\bar{x}$ =	57,88		***X*** *max* =	130,0
7	26,5	51,0	67,5		σ =	25,75		Δ***X/*** *M* =	17,3
8	27,5	53,5	68,5		α =	0,05			
9	31,0	55,0	70,0						
10	32,5	56,0	72,5		***X нач***	***X кон***	***n***	***n теор***	χ^2
11	34,0	56,0	77,5		9,00	26,30	3,00	5,50	1,14
12	36,0	56,5	81,0		26,30	43,60	12,00	8,98	1,02
13	36,5	57,5	82,5		43,60	60,90	16,00	12,85	0,77
14	39,0	58,0	90,0		60,90	78,20	11,00	11,92	0,07
15	40,0	59,0	96,0		78,20	95,50	3,00	7,15	2,41
16	41,0	59,0	101,5		95,50	112,80	2,00	2,78	0,22
17	42,5	60,0	117,5		112,80	130,10	3,00	0,82	5,75
18	43,0	61,0	127,5						
19	45,0	61,5	130,0					$\chi^2_{эмп}$ =	11,37
20	46,0	62,0						$\chi^2_{крит}$ =	9,49
22	*Эмпирическое значение* $\chi^2_{эмп} > \chi^2_{кр}$ *критического значения*								
23	*следовательно* ***гипотеза о нормальности X отклоняется***								

Рис. А – Скриншот расчетного листа MS Excel к примеру 1.1.

8. В ячейку H12 заносится формула =$F$4*(НОРМРАСП(F12; $F$6; $F$7; ИСТИНА) - НОРМРАСП(E12; $F$6; $F$7; ИСТИНА)) {=$F$4*(NORMDIST (F12; $F$6; $F$7; TRUE())-NORMDIST(E12;$F$6;$F$7;TRUE()))}, которая определяет для нормального закона распределения (заданы $\bar{x}$ и σ) теоретически ожидаемое число элементов выборки для второго диапазона данных (см. формулу для $n_i^{теор}$). Далее по содержимому H12 производится автозаполнение ячеек H12:H16 ожидаемого числа элементов выборки для всех остальных диапазонов (классов) кроме последнего.

9. В ячейку H17 заносится формула =$F$4* (1 - НОРМРАСП(E17;$F$6;$F$7; ИСТИНА)) {=$F$4*(1-NORMDIST (E17; $F$6;$F$7;TRUE()))}, которая определяет для нормального закона распределения теоретически ожидаемое число элементов выборки для отрезка от начала последнего диапазона данных до +∞.

10. В ячейку I11 заносится формула =(G11-H11)^2/H11, определяющая значение χ^2 (см. формулу $\chi^2_{экс}$) для пары частот первого интервала. Механизмом автозаполнения по содержимому I11 вводятся значения I12:I17 для остальных диапазонов.

11. В ячейке I19 формулой =СУММ(I11:I17) {=SUM(I11:I17)}

для числа классов I4-3 подсчитывается "полное" значение $\chi^2_{\text{экс}}$, в ячейке I20 формулой =ХИ2ОБР(F8;I4-3) {=CHIINV(F8;I4-3)} определяется критическое значение $\chi^2_{\text{крит}}$. Необходимо обратить внимание на подсчет числа степеней свободы, равного числу классов минус три, что обусловлено двухпараметричностью закона нормального распределения (см. п.4 общего алгоритма в разделе 1.2.1).

12. Сравнивая значения $\chi^2_{\text{экс}} > \chi^2_{\text{крит}}$ формулируется вывод:

нулевая гипотеза отклоняется, выборка не подчиняется нормальному закону распределения.